Analytical Tools for ORPS Data

Table of Contents

Introduction

Purpose of the Course

This class will provide you with an introduction to some of the basic concepts of the analysis of ORPS data and the use of various analytical tools in the evaluation of that data. You will learn the concept of process-focus and result-focus tools and their application to the analysis process. You will also receive information on using data from other sources to supplement ORPS data in order to further enhance the analysis process.

Disclaimer

You should be aware that most of the examples presented in this course are, of necessity, quite simplistic and are intended to illustrate concepts. In many cases, assumptions or adaptations have been made in order to illustrate the concepts and results may be based on incomplete data or analyses. Although the results of the examples may appear to indicate significant trends or issues, further investigation would be required in order to determine appropriate conclusions.

Objectives

Upon completion of this class, you will be able to perform the following activities:

- Describe the analysis process and how ORPS data can be used in that process

- Describe some of the data manipulation tools that can assist you in the analysis of ORPS data

- Explain the importance of normalizing data in the analysis process, identify some sources of normalizing data, and recognize the limitations of the normalizing data sources

- Identify the differences between process-focus and result-focus tools and describe how they can be used in the analysis of ORPS data

The Basics of Analysis

The Requirement for Data Analysis

For a number of years, the Department of Energy (DOE) has been collecting data in various databases reflecting facility operating experience. The reporting requirements for many of these databases are mandated by DOE order; for example, DOE Order 231.1, ***Environment, Safety, and Health Reporting***, and DOE Order 232.1A, ***Occurrence Reporting and Processing of Operations Information***.

In addition to collecting data, many of these orders, and the supporting requirements documents and manuals, also require an analysis of the data that is collected. This analysis may be required for lessons learned and accident prevention (e.g., DOE Order 232.1A) or accident investigation (e.g., DOE Order 225.1, ***Accident Investigations***).

The Stages of Data Analysis

Data analysis may be used in various stages in the life cycle of a system or process. If it is used during the design phase of a project, it will help to identify and correct problems prior to construction of a new facility or process. This early analysis helps designers identify and correct potential problems early when problems can be corrected at minimal cost. It also helps reduce the possibility of future accidents or incidents that may lead to unacceptable consequences during operation.

Early analysis of a facility must rely on design information and operating experience at other, similar, facilities. Once a facility has begun operation, the collection of actual operational data will provide additional sources of information. Evaluation of this data can provide a basis for additional changes that can further enhance facility safety and/or operability.

In spite of aggressive application of analytical techniques during the design and operation of a facility, accidents or other significant events, including near miss events, may still occur. In this case it will be necessary to use the analysis process to determine root causes and corrective actions in order to minimize the probability of a reoccurrence of the same or similar events.

Conducting an Analysis

Depending on the type of analysis being performed, you may find the analysis process either relatively straightforward or quite complex. In the event of an accident, the issue of basic concern is generally fairly well defined, although an adequate evaluation of causes may be complex. In other cases, there may be no single event of concern, and you may be looking for trends or other indicators that will warn of possible future concerns.

You will find that analysis of an issue is generally an iterative process. Results from one phase will provide you with data and scoping information for subsequent phases. You can continue with the analyses to more detailed levels until you identify root causes or potential failure mechanisms. The basic steps in the analysis process are as follows.

- Identify the issue(s) to be analyzed

  As previously noted, in some cases you may identify an issue as a result of an event at your own facility. This event may be an accident that results in unacceptable consequences, or it may only raise the possibility of future, more catastrophic, failures. Events at other similar facilities will also provide you with a basis for evaluation of your own facility. In these cases, the issue of interest is generally fairly well defined.

  In other cases, the analysis may be based on your desire to anticipate problems resulting from a new or changed process or system. In these cases, you may have little or no previous experience with the process in question and more extensive preliminary analysis may be required in order to define the issues.

- Identify the data that is available for, and applicable to, the desired analysis

  Some analysts make the mistake of relying on a single source of data in evaluating an issue or trying to predict future events. Instead, you should use multiple data sources where available in order to provide the broadest possible perspective on the issue.

- Gather and analyze the initial data

  The next step is to obtain raw or summarized data from the sources identified in the previous step. From these data, you will identify common characteristics, associations, trends, and findings that may lead to further investigation.

- Iterate the analysis to identify root causes and to formulate possible changes

  Based on your initial findings, you can then postulate additional issues for further study. You will need to continue this iteration until the basic causes or failure mechanisms are identified and changes can be proposed to minimize the probability and/or severity of future failures

- Analyze the effects of the recommended changes

Once you have completed the analysis and proposed changes, you must then repeat the analysis to the extent necessary to determine the effects of the changes and to prevent the introduction of new failure mechanisms as a result of the changes.

- Present the final conclusions and recommendations.

  In most cases you will be required to present formal documentation of your analysis.

Data Sources

Identifying the Sources of Data

One of the first tasks you must undertake in the analysis process is the identification of appropriate data sources. This section will provide you with an introduction to that task. For a more detailed discussion and an example of this process, you can refer to the document ***Practical Applications of TIS Series, Use of TIS to Target Needs for Improvement of Training Programs***.

Types of Data

As noted previously, a number of different databases are available within DOE that provide significant amounts of data regarding facility operation. Each of these databases individually contains important information that can be of great value when used in the analysis of ES&H issues. However, when data from multiple data sources is integrated into an analysis, the value of each source is greatly multiplied.

The use of multiple sources of data in performing analyses is desirable in order to provide the broadest possible perspective on a particular issue. The combination of views on an issue may help identify problems or solutions not apparent when using data from a single source.

The use of other, non DOE, data sources may also provide other insights into the characteristics of your own facility or process. In addition, you may have access to other, site specific, data sources such as system design documentation and test results, internal databases, etc., that are not available to outside analysts.

Available data can be broadly divided into three categories as follows.

- Actuarial Data - This is data that is based on actual operating experience at the same or comparable facility. This is the type of data that is most commonly considered when an analysis is initiated and may include raw or summarized data or both.

- Oversight Data - This data includes findings from previous audits. It may provide information such as specific design information, failure modes, or organizational deficiencies that are useful in determining root causes of events, potential initiating events, etc.

- Resource Data - This data consists of nonsite-specific material that can be used to determine applicable design criteria, obtain standard check lists, identify standard failure rates, etc.

Data Limitations

When using different data sources, you must remember that these sources may not include a consistent baseline or breakdown. For example, in the past, lack of coordination in the design and implementation of the various data systems has resulted in an inconsistency in the organizational structure used in the databases. Two of the most commonly used DOE databases, ORPS and CAIRS (the Computerized Accident/Incident Reporting System), vary greatly in their organizational structure. For example, substantial differences exist between the Field Organization and Contractor breakdowns in ORPS and CAIRS. Furthermore, ORPS subdivides Contractors to a Facility level, while CAIRS subdivides Contractors to an Operation Type (research, production, construction, etc.). ORPS identifies the DOE-HQ Program Office associated with an event, while CAIRS does not.

This inconsistency becomes critical when you compare data from one database with that from another, or when you use data from one database to normalize data from another. A number of attempts have been made in the past to provide organizational correlations between the various data sets, to allow better correlation of the data.

NOTE

One of the long term goals of the information system redesign is to standardize these organizational structures to provide better correlation between the various data sources.

Normalization of Data

The Need for Normalization

When using the data in ORPS in your analysis, you are limited by the fact that the database does not presently contain data elements for normalizing raw counts. In some cases, you may desire to compare the performance of your organization with the performance of other organizations or to evaluate the performance of your organization over time. Although direct trending over time or comparison of event counts between organizations can provide you with useful information, these direct comparisons can be misleading. The number of occurrences reported by an organization can be strongly influenced by the size or makeup of the organization, and trends over time are often dominated by a changing work force (contractor work force reductions), organizational realignment (reassignment of Program Office responsibilities), and changing reporting requirements (Order revisions). Therefore, if you are performing these types of comparisons using ORPS data, try to obtain data from other sources that will, at least to some extent, normalize the data being analyzed to a common baseline.

Additional discussion related to the need for the normalization of data is included in the DOE Office of Operating Experience Analysis and Feedback (OEAF) Technical Bulletin 96-1, ***Normalization of Data*** (http://tis.eh.doe.gov/web/oeaf/tools/oe_bull/oe_bull.html).

Sources of Normalizing Data

To the extent that a correlation can be established, data in ORPS may be normalized in a number of different ways. Some of the following have been suggested as possible types of normalizing data.

- The number of hours worked
- The number of employees
- The number of miles driven by fleet vehicles
- The hours of vehicle operation for air or rail vehicles
- The number of nuclear material movements
- The radioactive contamination area entries
- The volume of decontamination and decommissioning materials processed

Some of this data is readily available to you, while other types may only be available for your own facility or organization in local databases. Other types may not even be tracked or summarized, and may not be readily available.

CAIRS can provide a number of different sources of normalizing data, including the number of hours worked and the number of vehicle miles driven for various organizations, as well as property valuation information. Depending on the organization, organizational level, and type of activities being analyzed, you can make meaningful correlations of the data from the two databases. However, although normalization of ORPS data is possible using data from CAIRS, you must remember that, because of the differences in organization structure used in the two systems, direct association of data from the two systems is not always possible. In addition to the different organizational structures already mentioned, they also differ in reporting requirements and practices. Some organizations that report to CAIRS may not report to ORPS because of exemptions to the Order. For example, the Naval Reactors Program, which does report to CAIRS, is specifically exempted by DOE Order 232.1A from reporting to ORPS. Another example is the DOE Power Administrations, which report to CAIRS but not to ORPS. In order to obtain an accurate correlation of the two data sources, these differences must be recognized.

NOTE

Much of the correlation of data between data sets must be accomplished manually at the present time. Future enhancements to the data systems will help to automate many of these processes. One of the goals of re-engineering the various DOE ES&H data systems is to achieve consistency among the organizational structures within the various databases, thereby allowing more direct and accurate comparison of data from different sources.

An Example of Normalization

Because many of the analytical tools discussed in this class depend on the normalization of data, many of the examples have been chosen in areas where it is possible to obtain normalizing data from another source. The previously referenced OEAF Technical Bulletin suggested a possible correlation between construction related occurrences in ORPS and the number of work hours reported by construction organizations as reported in CAIRS. **Table 1** provides a correlation, at a Field Office level, of CAIRS organizations to ORPS organizations for the construction events considered in this study. This correlation includes field organizations having occurrence reports listing “Construction” as the activity category in ORPS and organizations listed as operation type “Cost Plus Construction” or “Lump Sum Construction” in CAIRS.

WARNING

Please note that this table does not include a complete mapping of all organizations at the Field organization level. Rather, it only includes those organizations that reported data in these categories.

Table 1. Mapping of ORPS and CAIRS organizational structure for construction occurrences.

ORPS ORGANIZATIONS		CAIRS ORGANIZATIONS	
Code	Description	Code	Description
ALO	Albuquerque	05	Albuquerque
CH	Chicago	10 28	Chicago National Renewable Energy Lab
HQ	Headquarters	20 95 96 97	Energy Technology Centers SSC Project Strategic Petroleum Reserves Yucca Mountain Project
ID	Idaho	30	Idaho
NVOO	Nevada	35	Nevada
OH	Ohio	45	Ohio
ORO	Oak Ridge	40	Oak Ridge
RFO	Rocky Flats	77	Rocky Flats
RL	Richland	75	Richland
SAN	San Francisco	80	Oakland
SR	Savannah River	85	Savannah River

Note: Construction activity was also reported in CAIRS by Pittsburgh Naval Reactors (Field Office code 60) and Schenectady Naval Reactors (Field Office code 90) for the time period covered by the ORPS data. Since these organizations are exempt from reporting to ORPS, this data was removed prior to calculating DOE total construction averages.

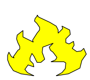
Obtaining and Formatting Data

The Basics

Once you have determined the source(s) of your data, you will then need to obtain the data from the data source and process it into a form that can be effectively used in your analysis or in a presentation to your management. Examples of such processing include the calculation of sums, averages, or confidence limits associated with a set of data or the normalization of the data. For limited data sets, you may perform these processes manually. In some cases, you will have access to specially developed database interfaces that will provide data in a summarized form that you can easily apply to a particular analytical tool, either manually or through automated processes. In other cases, summarized data may not be available in the particular form you are interested in. In these cases it may be necessary to enter raw data into a commercial software package such as Microsoft Excel or FoxPro and then perform your own manipulations of the data. The actual processes that you use will depend on the types of data available from the databases, and your own personal preferences or company standards. Some of the more common means of obtaining and processing ORPS data will be discussed in this section.

The HP Database and Visimage

The HP ORPS Search and Reports options provide a number of standard outputs that you can use as inputs to various analytical tools. Some of the more widely used output options are the User Defined Report, the Generic Lag Report, and the Distribution Reports. The User Defined Report and the Generic Lag Report provide information at a record level, while the Distribution Reports provide various summarizations for the selected records. Details of the content of these reports, as implemented on the ORPS GUI, are described in other sections of this manual.



WARNING

A problem has been identified with the HP software that can affect analysis results. This problem occurs when a search selection results in a set of occurrence reports that include roll-up reports where the number in the Number of Occurrences field is greater than 30.

Having multiple occurrences in a roll-up report is now permitted by DOE Order 232.1A, but the number of occurrences is limited to a maximum of 30. Prior to DOE Order 232.1, roll-up reports were not permitted. However, some organizations created the equivalent of a roll-up report by holding an occurrence report open and adding similar events as they occurred. This created some occurrence reports with greater than 30 occurrences. EH-33 is reviewing these reports and is updating the Number of Occurrences field to reflect the number of occurrences actually included in the reports. This sometimes results in the entry of numbers greater than 30.

You should be aware that if your search selection includes a roll-up report(s) with more than 30 occurrences, the number of occurrences for that report is counted as zero by HP ORPS. That means that, while the count of the number of reports (and graphs) resulting from a search will be accurate, the count of the number of occurrences will not. More importantly, these roll-up reports are also counted as zero occurrences in the distribution reports, which could cause significant errors. In this case, the defective roll-up reports are hidden within the other occurrence report totals and the error may not be detected.

As of 4/7/97, we have identified 13 roll-up reports with more than 30 occurrences identified in the Number of Occurrence field. These 13 reports contain 1,108 occurrences. In addition, the EH-33 task may create some additional roll-up reports with more than 30 occurrences. These reports can be identified by doing a search for Number of Occurrences > 30.

We believe that the problem is directly related to the method used in the count programs on the HP and that correcting the problem could be costly and could significantly impact the HP system performance. Since the use of the HP will eventually be phased out, the problem will probably not be corrected. This problem does not occur in the ORPS GUI or in reports generated with Visimage on the HP-3000 system.

The output options available from the HP for each of these reports include a display to the screen and corresponding printed records and electronic files. The electronic files are an ASCII text reproduction of the displayed and printed records. Once these output records are created, several options exist for transferring the data into one of the analytical tools. Many analysts will simply print out a hard copy of the report and manually enter the data into a spreadsheet or graphics package; however, other options are available to you.

When operating in a Windows environment and using Reflection for Windows, you can copy information directly from the screen and paste it into an application such as Microsoft Excel. This feature is useful for summarized data, such as the distribution reports, and tabular data, such as the generic lag report. However, this process has several limitations. Screen displays are generated one screen at a time, and longer lists of information will contain the prompts for continuing the display. Either this data must be captured one screen at a time, or the extraneous material must later be edited out. In addition, the data on the screen is delimited with spaces, and it may be necessary to change your default spreadsheet font in order for the columns to line up properly. Also, many of the distribution reports, in particular those that include subcategories such as the causes and nature of occurrence, are not formatted in a manner that lends itself to easy translation to the column format necessary for input to a spreadsheet.

In order to load data from the user defined report or other non summarized reports, it will probably be necessary for you to develop some type of special parsing program to extract the data and format it in a manner that can be used by the analytical tool. A number of organizations have developed programs of this type to load information from the user defined report or the complete occurrence report into a local database for additional processing.

You may also obtain direct access to the HP ORPS database through the use of a special database query tool named Visimage. Training in the use of Visimage is beyond the scope of this workshop, but those individuals who are trained can use it to easily obtain either raw or summarized data in a variety of formats that can be easily imported into another program for processing.

NOTE

The Visimage software access to the ORPS data will continue to be maintained while the database remains operational on the HP-3000 computer. Equivalent functionality will be provided with the ORPS GUI, through built-in functions within the interface and/or external access to the data.

The ORPS GUI

The new ORPS GUI is being designed with many features that will enhance the retrieval and use of ORPS data for analytical purposes. Many of these features have already been incorporated, while others are being planned for future release. Most of these features are discussed in detail in other sections of this manual. Some of the ones most applicable to data analysis are summarized as follows.

Search Enhancements

- Advanced Narrative Searches - The ORPS GUI provides greatly enhanced searching capabilities. The use of phrase searches and numerous special operators provide the capability to more accurately define a data set based on a narrative search.

- Record Screening Feature - The interface provides the ability to easily screen initial selections to identify and remove non-applicable records from the selection prior to saving a search profile or generating a report.

- Time Bounded Searches - Searches can be specified with date and time bounds applied to the data. Through the use of special time stamps applied to the data, a search can be specified as of a particular date and time, thus enabling search results to be reproduced or further refined without being affected by subsequent revisions to the data.

- Saved Search Profiles - An unlimited number of search profiles can be saved, and saved profiles can be edited.

Report Enhancements

- Graphics Availability - Direct generation of graphic reports (bar charts) is available from the interface.

- Tab Delimited Data - Data on distribution is available from the distribution and graphic reports. This data can easily be copied and pasted into another application.

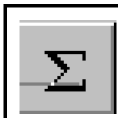
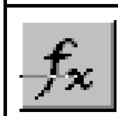
Additional enhancements that will continue to improve data access are also planned for future releases of the application.

Using a Spreadsheet Software Package

If you intend to do extensive analysis of ORPS data, particularly for applications such as trending of data, you may benefit from learning some of the basics of a software package. Although detailed instruction in the use of a spreadsheet is beyond the scope of this course, we will demonstrate some of the basic capabilities that can be easily used to aid in processing data. Since Microsoft Excel is the most widely used spreadsheet software in the DOE complex, we will use it to demonstrate a few analysis techniques.

Sums and Functions

Excel contains a number of built in functions that can help you process your data for further use in your analysis. Most of these are available from two buttons found on the tool bar.

	The sum function is useful for quickly calculating sums of rows or columns of data. To use the sum function, simply highlight the series of data that you wish to sum and click on the **SUM** button in the Tool bar. The sum will automatically be calculated and displayed at the end of the series.
	Excel also contains a large number of additional functions that can be accessed by clicking on the **FUNCTION** button on the Toolbar. This will display the Function dialog box, shown in **Figure 1**.

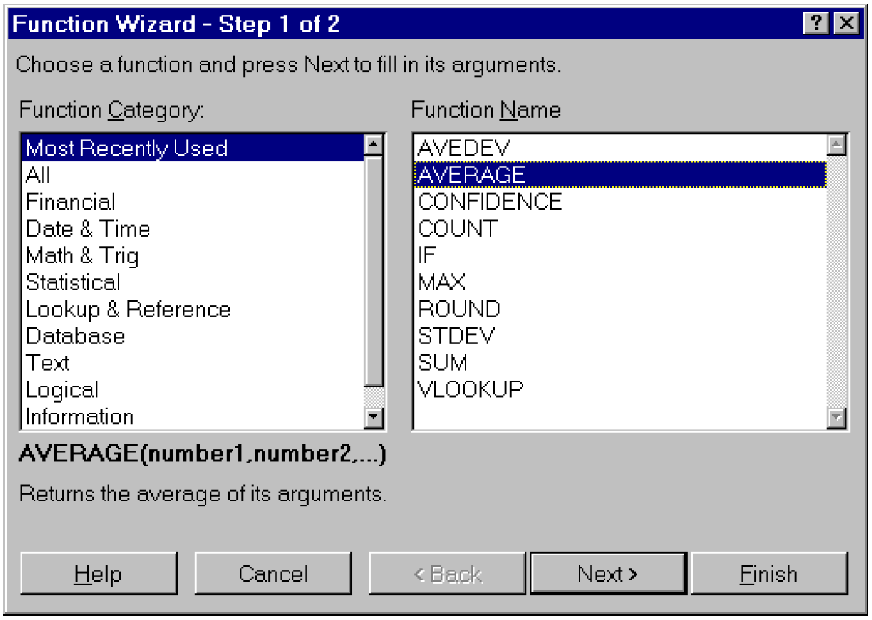


Figure 1 - The **FUNCTION WIZARD** dialog box.

From the **FUNCTION WIZARD** dialog box, you can select a function to apply to your data. For this example, we will select the **AVERAGE** function. By highlighting the **AVERAGE** function and clicking the **NEXT** button, the **AVERAGE** dialog box is displayed (**Figure 2**). Enter the range of data that you want to average either by typing in the dialog box or by highlighting the range with the mouse. When you click on the **FINISH** button, the average value of the data will be calculated and displayed at the current cursor position.

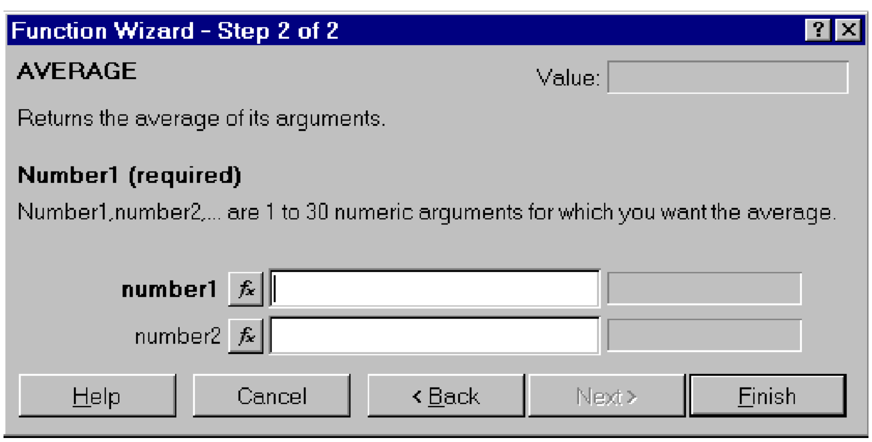


Figure 2 - The **AVERAGE** dialog box.

The Data Menu

By clicking on **DATA** in the Excel menu bar, the menu in **Figure 3** is displayed. This menu contains a number of items that allow you to further manipulate your data. The **Text to Columns** function, described in the ORPS GUI Basic Report Techniques section of this manual, is found here. We will discuss two additional features, the **Sort** function and the **Pivot Table**.

Figure 3 - The Excel **DATA** menu.

Sorting Data

In many cases, the data that you have loaded into your Excel spreadsheet may be in random order. In order to use the **Sum** and **Average** functions described previously, the data needs to be ordered consistent with the desired result. This can be easily accomplished with the **Sort** feature. To sort a set of data, highlight the data to be sorted, then open the **DATA** menu and select **SORT**. The dialog box in **Figure 4** is presented. This dialog box allows you to specify up to three sort levels, and each level may be sorted in increasing or decreasing order. After making the appropriate selections, click the **OK** button to complete the sort.

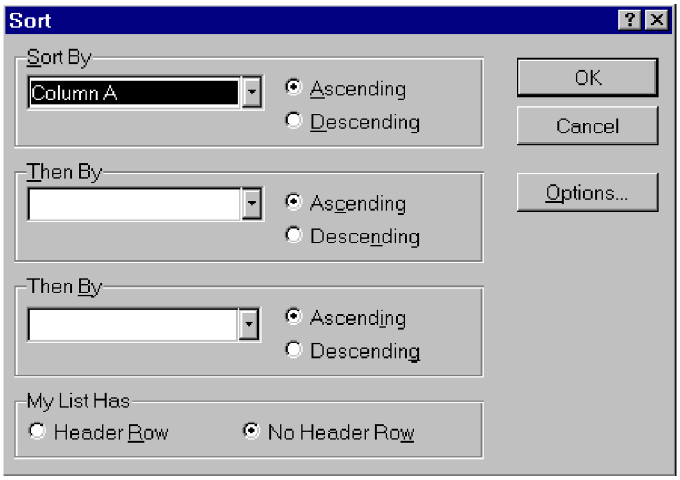


Figure 4 - The **SORT** dialog box.

The Pivot Table

The Excel **Pivot Table** provides a quick and simple means of summarizing data from a spreadsheet. A simple application of the Pivot Table will be demonstrated using a file containing data for each record in ORPS having a categorization date in March 1997. The file was created using the Visimage software and includes the following fields for each of the records.

- Field Office
- Contractor
- Facility ID
- Year
- Sequence Number
- Report Type
- Occurrence Category
- Direct Cause
- Root Cause

Once the file has been loaded into Excel, the **Pivot Table Wizard** is started by selecting **Pivot Table** from the **Data** menu. The dialog box in **Figure 5** will appear.

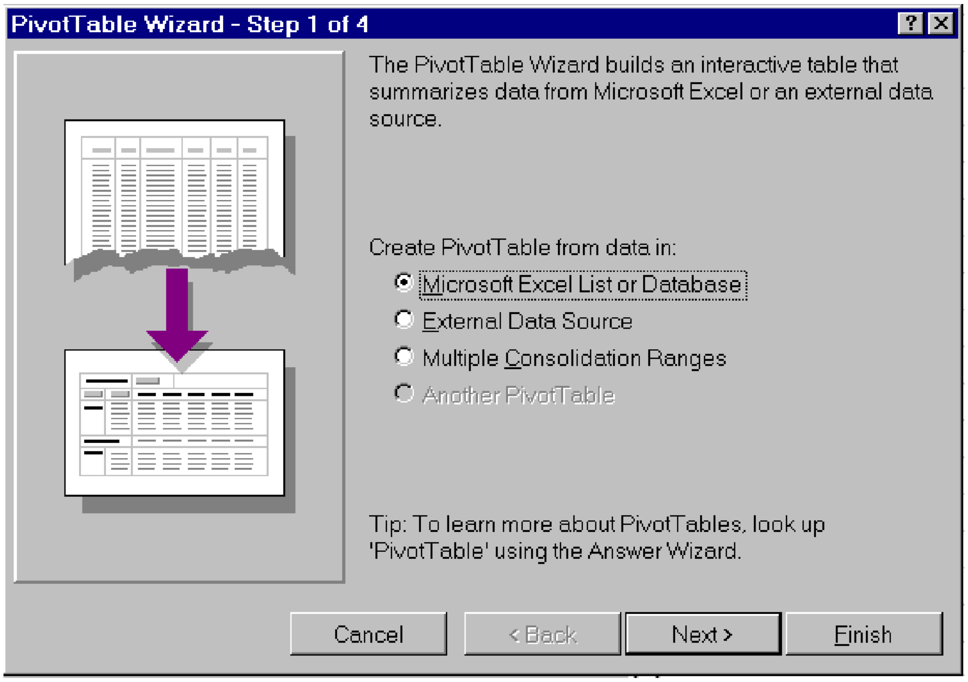


Figure 5 - The **PIVOT TABLE WIZARD** dialog box for selecting data type.

Select **MICROSOFT EXCEL LIST OR DATABASE** and click **NEXT** to proceed to the dialog box in **Figure 6**.

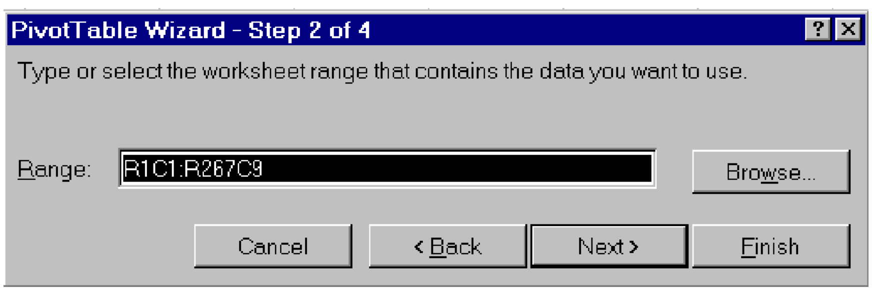


Figure 6 - The **PIVOT TABLE WIZARD** dialog box for selecting data range.

In this box, you select the range that you want to include in the distribution. This can be done either by typing in the range or by selecting the range from the spreadsheet using the mouse. Once the output range is specified, click on **NEXT** to continue to the dialog box in **Figure 7**.

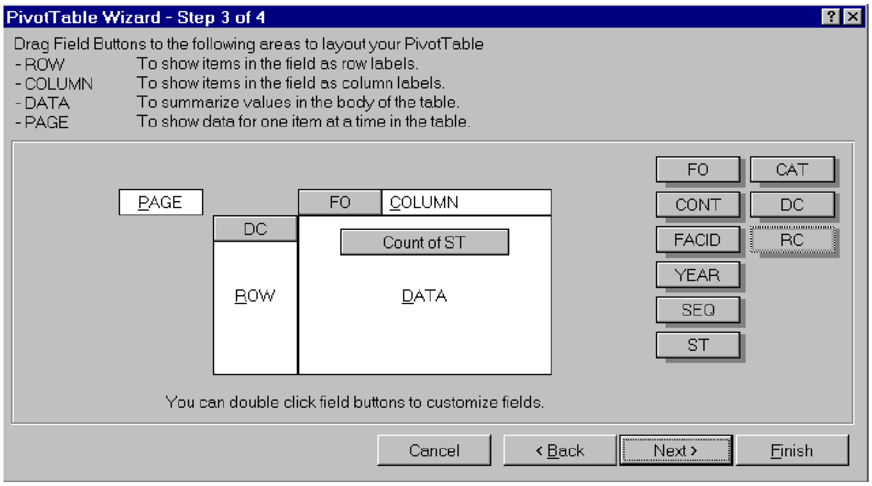


Figure 7 - The **PIVOT TABLE WIZARD** layout dialog box.

Select the items to distribute by dragging the item to the appropriate location on the layout. For this example, we will list the **Field Offices** in columns and the **Direct Causes** in rows. The final step is to select an item to total. In this case we will just be counting records, so any field that will occur in each record can be selected. In this case we have selected **Report Type**. Other totaling options are also available. If we had included the number of occurrences as a field in the file, we could sum the field to obtain the number of occurrences. Once the layout is complete, click on **NEXT** to continue to the dialog box in **Figure 8**. In this box, enter the cell location for the starting cell of the output. Again, the cell location can be typed or entered with the mouse. Finally, click **FINISH**, and the distribution will be created as shown below.

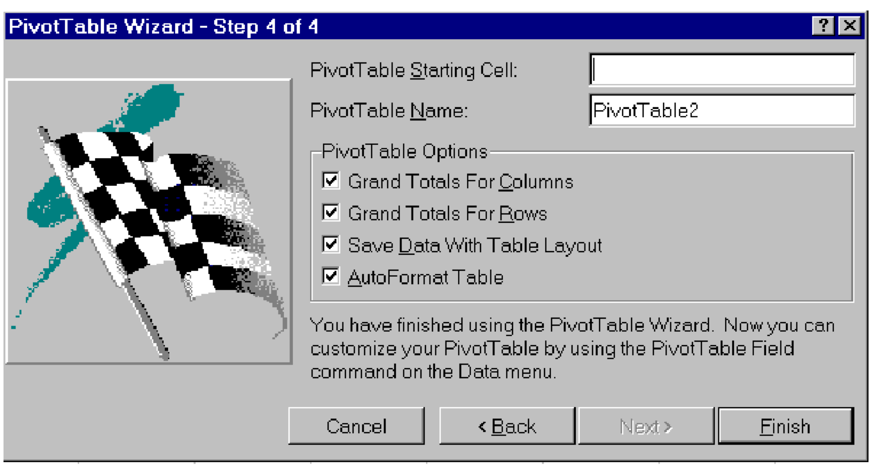


Figure 8 - The **PIVOT TABLE WIZARD** dialog box for selecting starting cell.

Count of ST	FO										
DC	ALO	CH	HQ	ID	NVOO	OH	ORO	RFO	RL	SAN	SR	Grand Total
1A	4	2	1	0	0	0	0	9	0	0	1	17
1B	1	0	2	0	0	0	0	0	0	0	0	3
1E	0	0	0	0	0	0	0	1	0	0	0	1
1F	0	0	0	0	0	0	0	0	2	0	0	2
2A	0	0	0	0	1	0	0	0	2	1	0	4
3A	0	1	0	2	0	0	0	1	1	0	0	5
3B	0	0	0	0	0	0	0	1	0	1	3	5
3D	1	0	0	0	0	1	0	1	1	0	0	4
6A	1	0	0	0	0	0	0	0	0	0	0	1
6B	0	0	0	0	0	0	0	0	0	0	1	1
7A	0	0	0	0	2	0	0	0	0	1	0	3
7C	1	0	0	0	0	0	0	0	0	0	0	1
8A	0	0	0	0	0	0	1	0	2	0	0	3
8B	1	0	0	0	0	0	0	0	0	0	0	1
(blank)	35	16	2	13	3	12	28	13	44	8	41	215
Grand Total	44	19	5	15	6	13	29	26	52	11	46	266

The ORPS Excel Toolkit

The DOE Office of Operating Experience Analysis and Feedback (OEAF) has developed an ORPS Parser and other related software that uses Excel for analyzing ORPS data. In addition to automatically loading and parsing data from the distribution reports, the software also contains several statistical tools. The software is available for download from the OEAF website and contains detailed instructions for installation and operation as well as several sample data files. In this section we will illustrate the operation of the parser portion of the software.

Currently, the parser only operates on ORPS distribution reports that are downloaded from the HP ORPS. The parser will presently work with all distribution reports created with HP ORPS. However, it is not necessary to include all of the distributions in the downloaded file, and the order of the distributions is not important. For this example we have created a file named DISTDATA.TXT that includes a distribution of 1997 reports by Direct Cause, Facility Function, Discovery Year/Quarter, and Program Office.

Once the parser software has been installed, a new menu item (**ORPS**) will appear in the menu bar (**Figure 9**). A new ORPS tool bar will also be available, as shown in **Figure 10**.

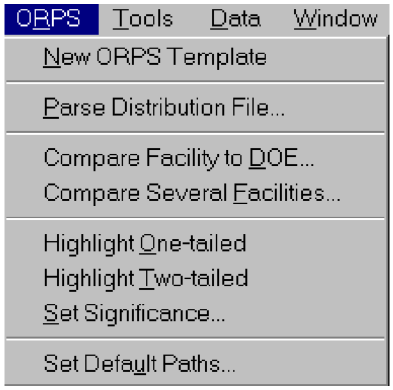


Figure 9 - The **ORPS Toolkit** Menu.

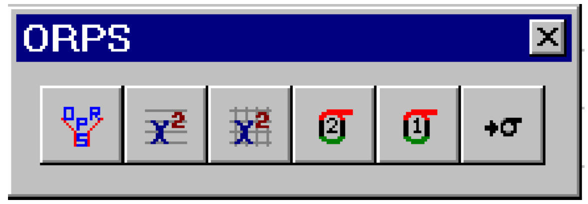


Figure 10 - The **ORPS Toolkit** Button Bar.

To load and parse the distribution data, select **NEW ORPS TEMPLATE** from the **ORPS** Menu. Then select **PARSE DISTRIBUTION FILE** from the **ORPS** Menu or click on the left button on the **ORPS Tool Bar**. The following dialog box will appear.

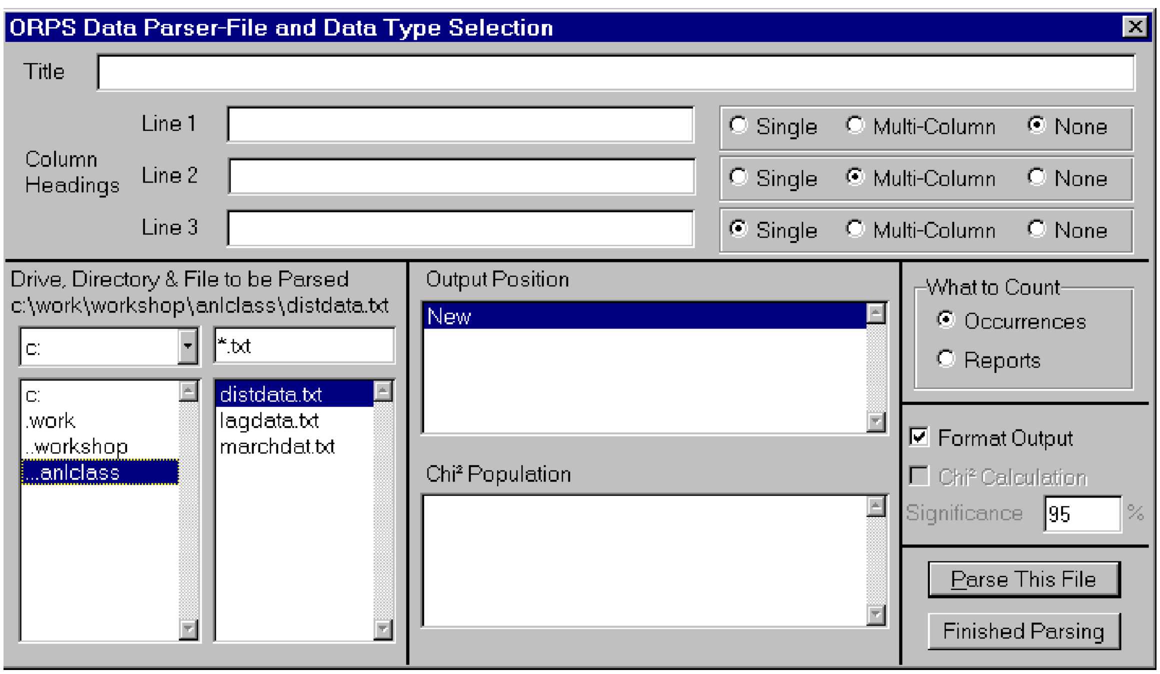


Figure 11 - The **ORPS Data Parser** dialog box.

Using the dialog box, select the distribution file to be loaded. A report title can be entered if desired, and the program can be instructed to count either occurrences or reports. Once the proper selections have been made, click the **PARSE THIS FILE** button to continue The data will be loaded directly into Excel and will be available for further processing or charting.

Types of Analytical Tools

The Basics of Analytical Tools

In this section we will provide you with an overview of some of the specific analytical tools used in data analysis and discuss their application to ORPS. The majority of the analytical tools presented in this document are tools that are in common use throughout a wide range of disciplines. Consequently, a number of published sources describe in varying degrees of detail the fundamentals of, and the theory behind, many of the tools. Therefore, the descriptions contained in this document will be limited and the discussion will focus primarily on the application of the tools.

NOTE

The TRADE publication, ***How To Measure Performance, A Handbook of Techniques and Tools***, contains descriptions of a number of analytical tools. A future TRAC publication, ***Practical Applications of TIS Series, Tools for the Evaluation of Data in an Integrated Environment*** will also describe a number of tools and provide detailed references to sources of additional information.

Analytical tools can be broadly grouped into two types, process focus tools and result focus tools. However, the distinction between the two types of tools is not always well defined. Some result focus tools may be used as an integral part of a process focus tool. The tool is used to display information at some intermediate point in an analysis so that a decision can be made as to the direction that further analysis should proceed.

Process Focus Tools

Process focus tools scope a system or process and are used for analyzing data and developing relevant conclusions. They generally consist of a structured methodology, ranging from very simple to very complex, that is used to identify and resolve problems and issues. A number of process focus tools are described below with sample applications.

Barrier/Control Analysis

Barrier Analysis, also known as Control Analysis, is a tool to evaluate barriers (either physical or administrative) to prevent the unwanted flow of energy to targets (personnel or equipment) to prevent an accident or incident. Barrier analysis is useful as a qualitative tool for systems analysis or safety reviews (accident prevention) or for post-accident failure analysis. The actual implementation of Barrier Analysis can take a variety of forms, including graphical, tabular, checklist, etc.

Step-by-Step
TUTORIAL

Steps in Performing a Barrier Analysis

1. Identify potential (or actual) energy sources and their possible targets.

2. Identify the barriers that exist to prevent the flow of energy to the target. In the case of a post accident analysis, identify the barrier(s) that actually failed or that prevented the event from being more severe.

3. Identify the failure mechanisms (postulated or actual) for each barrier, and the degree of redundancy of the barriers.

4. Identify the effects of failure of the barriers.

Although Barrier Analysis has primarily been considered a qualitative tool, the U.S. Department of Energy, Office of Operating Experience Analysis (EH-33), has developed a variation that provides quantitative results which can be used as input to risk assessments. This variation uses an event tree format to display combinations of barriers and the overall consequences of their failures. Probabilities can be assigned to each of the branches based on operating experience, other actuarial data, etc. The probabilities can be combined to estimate the probability of occurrence of a combination of failures. A sample is shown in **Figure 12**.

Barriers to Injury

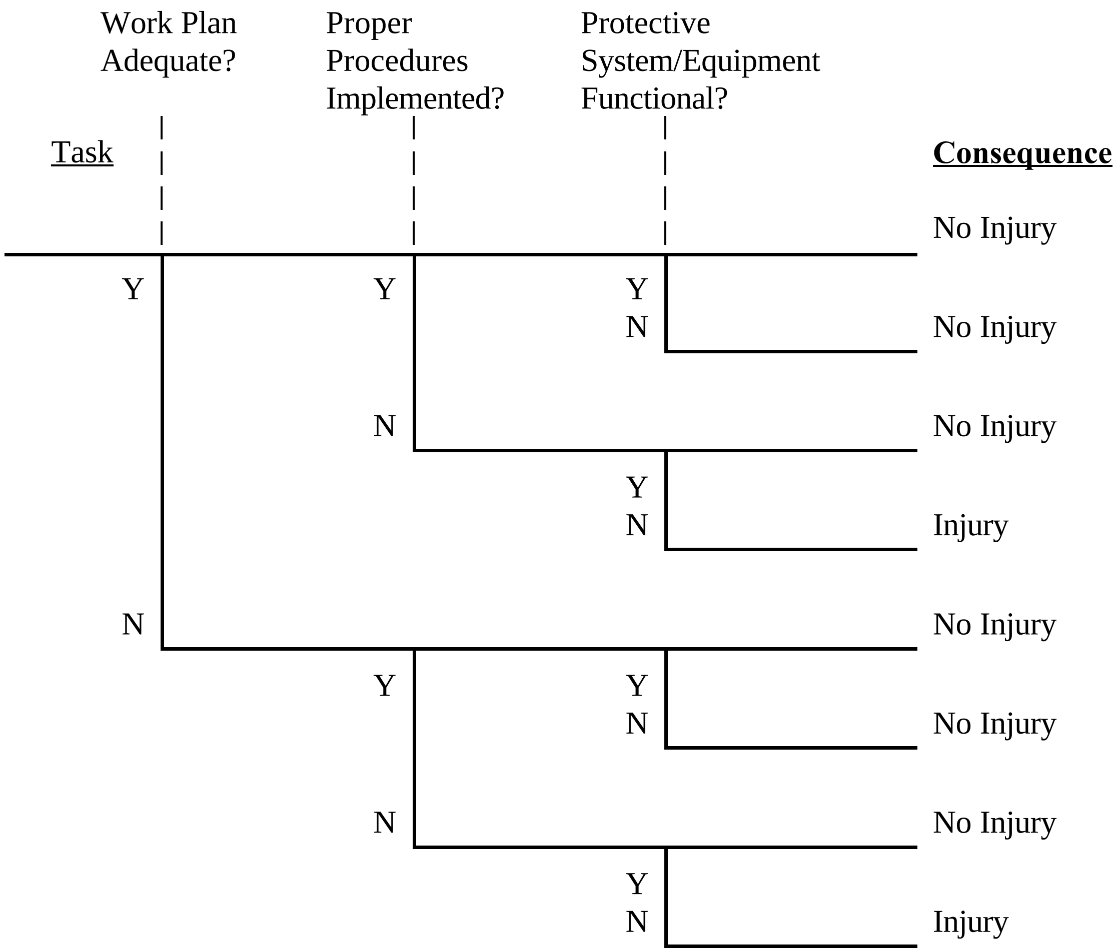


Figure 12 - Sample Barrier Analysis

Change Analysis

Change analysis examines the potential effects of modifications from a starting point or baseline by systematically hypothesizing worst-case effects from each modification from that baseline. It is useful as an accident investigation tool as well as a tool for investigating the effects of proposed changes to a system. The analysis is carried out by comparing the accident situation (actual or postulated) with the prior, ideal, or accident free situation. Differences are noted and the effects of the differences are evaluated. Factors are considered in terms of what, when, where, who, managerial controls, and others as appropriate.

Step-by-Step

TUTORIAL

Steps in Performing a Change Analysis

1. Identify the accident situation.

2. Identify a comparable accident free situation.

3. Compare the two situations and note the differences.

4. Analyze the differences for their effects on the accident.

An ORPS report from Wackenhut Services, Inc. at the Savannah River Site (SR--WSIS-SECFOR-1995-0002) describes a fatality resulting from a fall during rappel training. The accident resulted in a Type A investigation. A portion of the Change Analysis for that event as documented in the investigation report is shown in **Figure 13**.

Accident Situation	Prior, Ideal, or Accident-Free Situation	Difference	Evaluation of Effect
Safety rail gates closed	Safety rail gates open	With gates open, the rappel is performed from the scuff pad on the tower floor. With the gates closed, the rappel was done with rope on top of the safety rail.	The decision was made to Buddy Rappel with the rope on top of the safety rail and with the gate closed.
Scuff pads were not used on top of safety rails	Scuff pads on edge of tower floor	Rope contacted small radius edge of lock-pin housing.	The rappel rope received no protection from the small radius edge of the lock-pin housing.
Lesson plans not required for SWAT competition training	Lesson plans for training	Training is informal, not reviewed by Safety, and not approved by management.	SWAT competition training is informal, planning is insufficient, and hazardous activities are included.
Inspection of safety rails not performed - procedures not followed	Rappel Procedure 3-6601 requires a visual inspection of the tower prior to each iteration of rappelling	Safety rails were not inspected for rope pathway hazards - a small radius edge.	Lack of inspection led to hazardous rope pathway during SWAT competition.

Figure 13 - Sample Change Analysis

Check Sheet

The check sheet uses a list of specific items to identify known types of hazards, design deficiencies, and potential accident situations associated with common equipment and operations or to otherwise guide the analysis of a particular situation.

Step-by-Step
TUTORIAL

Steps in Utilizing a Check Sheet

1. Select or develop an appropriate check sheet.

2. Perform the review.

3. Document the results.

The Los Alamos National Laboratory Occurrence Investigation Group has developed a checklist (**Figure 14**) that is used to determine the extent of required analysis for an occurrence. The checklist is jointly completed by the Occurrence Investigation Group and the Facility Manager (Initial Evaluation). The checklist is later validated by the Occurrence Investigation Group (Final Evaluation). If answers to all questions are “No,” the Occurrence Report requires minimal analysis. A “Yes” answer to any question indicates that the occurrence should be subjected to a graded investigation and causal factor analysis.

	Initial		Final	
1) Was this an emergency or unusual occurrence?	Yes	No	Yes	No
2) Was there a near miss to an emergency or unusual occurrence?	Yes	No	Yes	No
3) Does the occurrence involve a compromise of personal safety?	Yes	No	Yes	No
4) Does the occurrence involve safety-related equipment?	Yes	No	Yes	No
5) Does the occurrence involve a significant deviation from the authorization basis (OSRs, TSRs, SARs,) or violate other operating safety limitations?	Yes	No	Yes	No
6) Was there significant personal injury or illness?	Yes	No	Yes	No
7) Was there significant environmental impact?	Yes	No	Yes	No
8) Was there reportable personal radiological contamination?	Yes	No	Yes	No
9) Was there radiological contamination off-site or in a publically accessible area on-site?	Yes	No	Yes	No
10) Is recurrence (prior or future) an issue?	Yes	No	Yes	No
11) Was there a strong indication of systematic or programmatic issues or of Laboratory-wide implications?	Yes	No	Yes	No

Figure 14 - Sample Check Sheet

Economic Matrix

The economic matrix is a procedure for prioritizing issues based on a combination of two parameters. The first parameter is a measure of total loss, which gives an indication of the economic impact to DOE for each issue. The second parameter is a loss rate, which recognizes the fact that larger organizations will generally have higher losses than smaller organizations. Issues are placed on a matrix, shown in **Figure 15**, based on rankings within the two parameters. Issue priorities are assigned based on the position within the matrix.

Primary	High	Medium Priority	High Priority	Very High Priority
	Medium	Low Priority	Medium Priority	High Priority
	Low	Very Low Priority	Low Priority	Medium Priority
		Low	Medium	High
		Secondary Indicator		

Figure 15 - The Economic Matrix

The Economic Matrix can be used to prioritize issues and focus resources where they will have the most potential for return. It can be used at any level that will support the ranking of issues, i.e., for which both loss and rate data exist. Rates do not have to be only time based; meaningful results have been obtained when evaluating severity based rates.

Step-by-Step

TUTORIAL

Steps in Creating an Economic Matrix

1. Identify the parameters to be used in the matrix. One parameter should be a direct or indirect measure of total loss of each issue. The other parameter should be a rate that is based on some measure of activity level.

2. Identify the ranges of the parameters to be used in ranking.

3. Rank each issue within each parameter.

4. Plot each issue on the matrix. Adjust the parameters as necessary to achieve a good distribution.

5. Select the highest priority issues for further evaluation.

In the following example, the economic matrix is used to analyze all of the 1995 occurrences that specified Construction as the Activity Category for the occurrence.

1. Identification of parameters.

Loss parameter - The majority of ORPS reports do not identify an actual loss value, so some method of deriving equivalent losses must be defined. For purposes of this example, a derived cost based on the number and category of occurrences was used. The cost assigned to each occurrence was the value used for determining reportability under the value based reporting criteria, i.e., $10,000 for an Off-normal occurrence and $1,000,000 for an Unusual occurrence. The single Emergency occurrence included in the data set was also assigned a value of $1,000,000. Where multiple Off-normal occurrences were included in a single occurrence report, $10,000 was assigned for each occurrence. These values were used for illustration purposes only. In an actual application, values may be assigned that are more representative of actual costs associated with the specific type of occurrences being evaluated.

Rate parameter - The number of occurrences was normalized to a rate per 1,000,000 construction hours worked. The number of hours was obtained from the CAIRS data base by summing the work hours reported by organizations categorized as Cost Plus Construction or Lump Sum Construction organizations.

2. Identification of ranges.

   Loss parameter - An upper value for the low range for total loss equal to the point at which 80 percent of the loss has been accounted for when using a Pareto analysis of the loss has been found to be a good starting point. Thus 80 percent of the loss will fall in either the medium or high range. For this particular data set, this value is $1,300,000. (See the Pareto analysis example under the Result Focus Tools section.) The upper value for the medium range for total loss can be initially set to some multiple of the lower value or to some other discrete value. This value may require adjustment to provide a good distribution within the matrix. For this particular data set, a value of $2,000,000 gave a good distribution for the final matrix.

   Rate parameter - A good starting point for an upper value for the low range for loss rate is the average loss rate for the entire population being. An organization having a rate less than the average is generally considered to be of low interest. For this particular data set, this value is 9.86 occurrences per 1,000,000 construction work hours. The upper value for the medium range is normally set to some multiple (typically 2 or 3) of the average. For this example, a value of 2 times the average rate, or 19.72 occurrences per 1,000,000 construction hours gave a good distribution for the final matrix.

The distribution of Field organizations using the ranges defined above is shown in **Figure 16**. Based on the screening, the Chicago Operations Office receives a very high ranking and would be the highest priority for further evaluation. This is based on a combination of a high occurrence cost ($2,240,000 based on 2 unusual occurrences and 24 off-normal occurrences) and high occurrence rate (27 occurrences per 1,000,000 construction hours). Headquarters and the Idaho Operations Office received high rankings and would be the next priority. If resources are available, medium ranked Field organizations (Albuquerque, Ohio, and Richland) could be evaluated.

Occurrence	>$2M	RL ALO	HQ	CH
	≤$2M ≥$1.3M	SR		ID
	<$1.3M	SAN NVOO	RFO ORO	OH
		<2.0	≤ 4.0 ≥2.0	>4.0
		Occurrence Rate		

Figure 16 - Sample Economic Matrix

Events and Causal Factors Analysis

Events and causal factors analysis is a methodology that uses a block diagram to depict cause and effect. It is used to develop root causes associated with an event.

Events and causal factors analysis lends itself well to a team-based approach to causal factor determination. A team of experts, with representatives from several different disciplines, can work together to systematically track causes and effects to increasingly lower levels until the actual root cause(s) is located.

Step-by-Step
TUTORIAL

Steps in Performing an Events and Causal Factors Analysis

1. Arrange events chronologically from left to right.

2. Enclose events in rectangles and conditions in ovals.

3. Connect events by solid arrows.

4. Connect conditions to each other and to events with dashed arrows.

5. Presumptive events or conditions should be enclosed in dashed line rectangles or ovals.

6. Secondary events, contributing factors, and systemic factors should be on horizontal lines at different levels above or below the primary sequence.

7. To specify individuals, break out each person on a separate horizontal line.

8. Events should track in logical progression from the beginning to the end of the accident sequence.

An ORPS report from the National Renewable Energy Laboratory (CH-NA-NREL-NREL-1995-0012) discusses the evacuation of a building as a result of a natural gas leak. Although no significant consequences occurred from the event, investigation of the event showed numerous failures in equipment and procedures. A small portion of an events and causal factors diagram for this occurrence is shown in **Figure 17**.

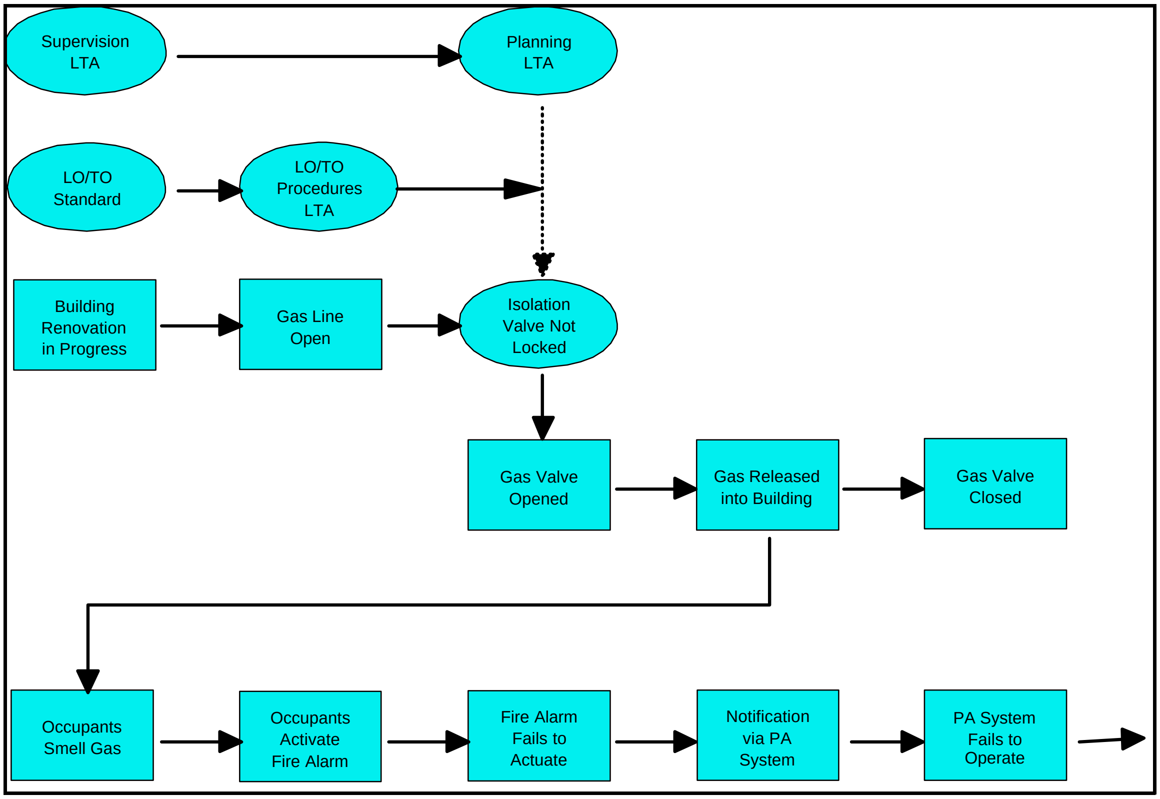


Figure 17 - Sample Events and Causal Factors Analysis

Extreme Value Analysis

Extreme value analysis, or extreme value projection, is a risk projection technique that can be used to predict, or provide information about, losses that are greater than any losses that have occurred to date. It can be used to calculate probabilities or return periods associated with certain infrequent, high loss events.

Step-by-Step
TUTORIAL

Steps in Creating an Extreme Value Projection

1. Select the type of loss to be studied.

2. Select the time span desired for the projection and the time intervals within the span. Ideally, a time span should have between five and twenty time periods.

3. Collect the worst case loss data for each time period.

4. Rank the loss data from lowest to highest.

5. Calculate the cumulative probability for each item. This can be approximated by dividing the item number from the order ranked list by the total number of items plus one (P = i/n+1).

6. Select the scale for the loss axis. The scale should be such that the highest value of loss during the time period plots between one-half and three quarters of the way up the y-axis.

7. Plot the points on the special extreme value paper by using the cumulative probability for the x-axis and the value of the maximum loss for the y-axis.

8. Draw a best fit straight line through the points.

NOTE

Extreme value paper comes in two different types. In one type, the loss scale is linear; in the other type it is logrithmic. If a good straight line approximation cannot be obtained with one type, try the second type.

Property loss data from CAIRS for the time period 1976 to 1995 was evaluated to determine if the high loss events seen during that time were expected losses or were less probable outliers. The worst-case losses for each year, ranked in order of loss, and the approximate cumulative probability are as follows.

Rank	Cumulative Probability	Year	Loss
1	0.048	1989	10000
2	0.095	1994	14000
3	0.143	1982	25152
4	0.190	1993	25238
5	0.238	1977	30000
6	0.286	1991	32000
7	0.333	1976	45000
8	0.381	1985	47000
9	0.429	1988	47606
10	0.476	1986	49500
11	0.524	1980	50000
12	0.571	1990	54000
13	0.619	1981	65489
14	0.667	1979	70000
15	0.714	1995	71074
16	0.762	1984	87067
17	0.810	1987	174400
18	0.857	1983	240000
19	0.905	1978	692136
20	0.952	1992	1000000

(Note that the CAIRS system has a built in utility for identifying worst-case events, ranking the losses, and calculating a more exact cumulative probability.)

The resulting data points can then be plotted on extreme value paper as shown in **Figure 18**.

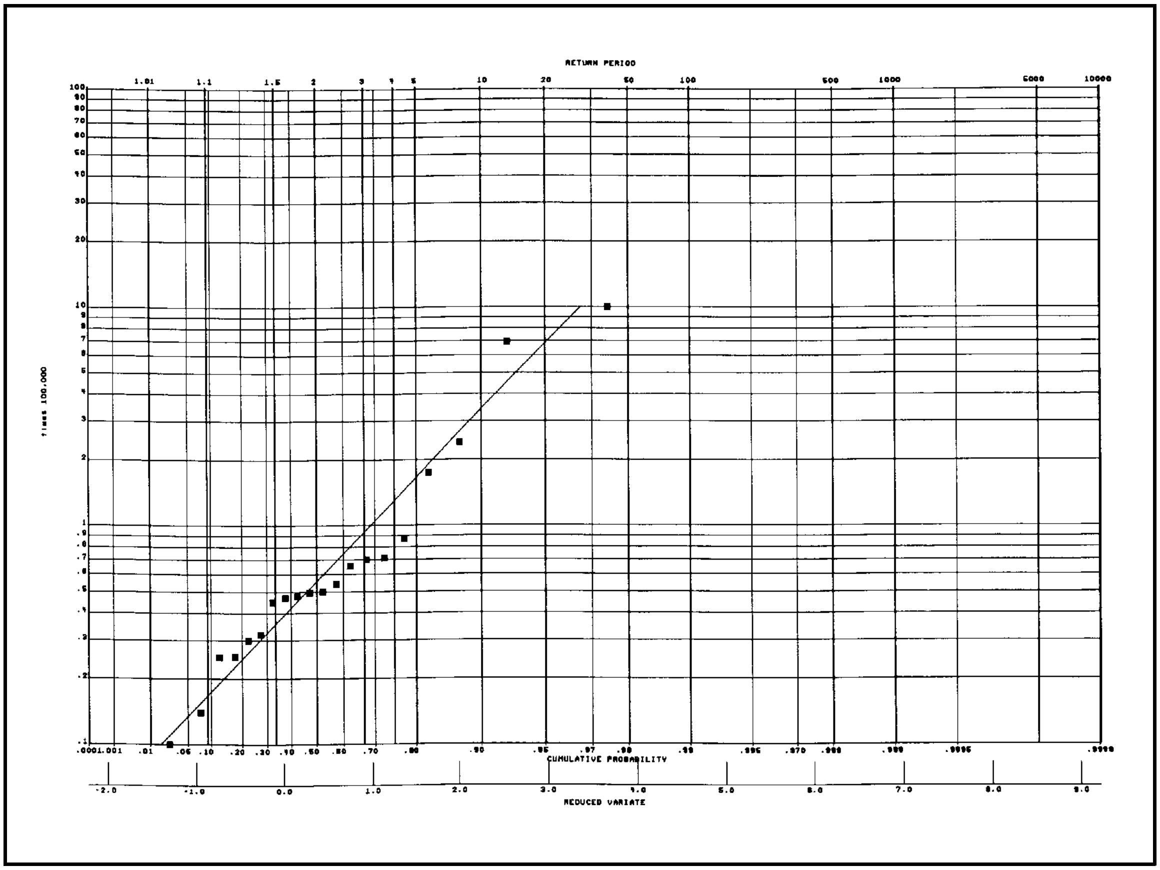


Figure 18 - Sample Extreme Value Projection. Loss value ($ in 100,000) is plotted on the Y-axis, Cumulative Probability of that loss is plotted on the X-axis.

Result Focus Tools

Result focus tools are used to display the results or conclusions of a study. In this section we will describe a number of these tools and provide you with example applications. The examples will primarily use data from ORPS.

Many modern PC packages provide a charting feature that will simplify the creation of most of these tools. You can simply enter data into a spreadsheet (Excel) or table (Freelance) and the software will create the necessary chart structure and plot the data points. The charts in this section were created using Lotus Freelance. Once the charts have been created, they can be easily copied and pasted into another document as was done for this workbook.

Bar Chart

Bar charts are used to provide easy visualization of data distributions or trends. They can be used to provide comparisons of various issues or to show trends in an issue over time. The procedure used to create a bar chart is as follows.

Step-by-Step TUTORIAL

Steps in Creating a Bar Chart

1. Assemble the data to be analyzed.

2. Determine appropriate scales for the charts.

3. Plot the data on the chart.

4. Analyze the chart for trends.

The bar chart in **Figure 19** shows the rate of construction occurrences per 200,000 construction work hours for contractors reporting to the Chicago Operations Office. The Chicago Operations Office was shown with the Economic Matrix to be a high priority for further evaluation. The bar chart shows BNL to be the highest contributor to the overall rate.

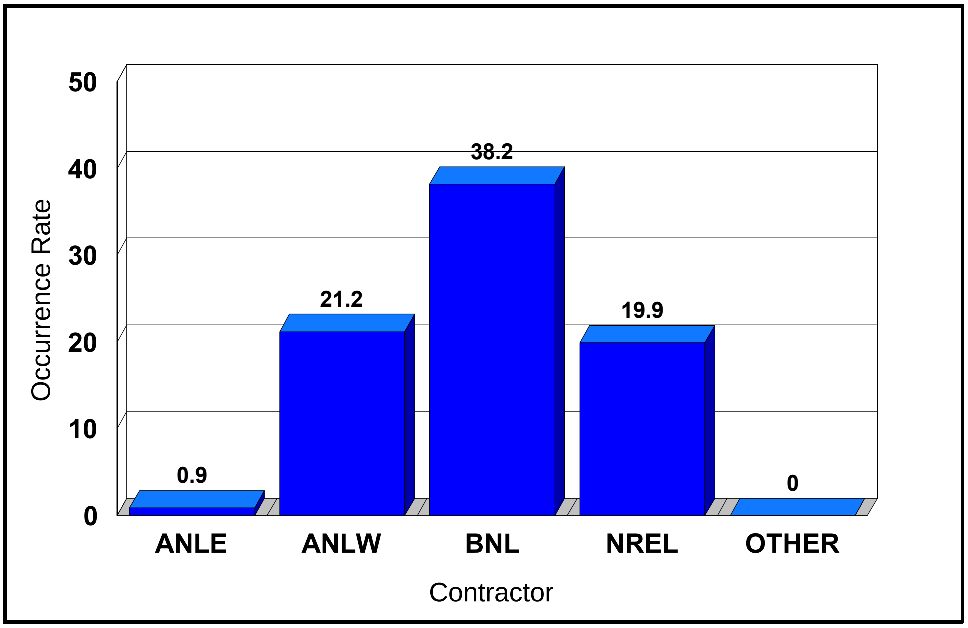


Figure 19 - Sample Bar Chart

Control Chart

A Control Chart is a variation of the line chart that contains control limits. The Control Chart provides a visual indication of trends in the data as well as outliers from what would be considered a statistically normal distribution, e.g., several continuous data points that show a continuous upward or downward trend or individual data points that fall outside of the control limits.

Step-by-Step
TUTORIAL

Steps in Creating a Control Chart

1. Assemble the data to be analyzed.

2. Calculate the average of the data points and the control limits.

3. Plot the data points on the chart.

4. Add lines representing the average and the control limits.

5. Evaluate data points that fall outside the control limits to determine possible causes.

The example in **Figure 20** shows the number of construction related occurrences reported by the Chicago Operations Office distributed by year and quarter from 1991 through 1995. An upper control limit equal to three standard deviations from the average is shown. The calculated three standard deviation lower control limit is below the zero point on the axis.

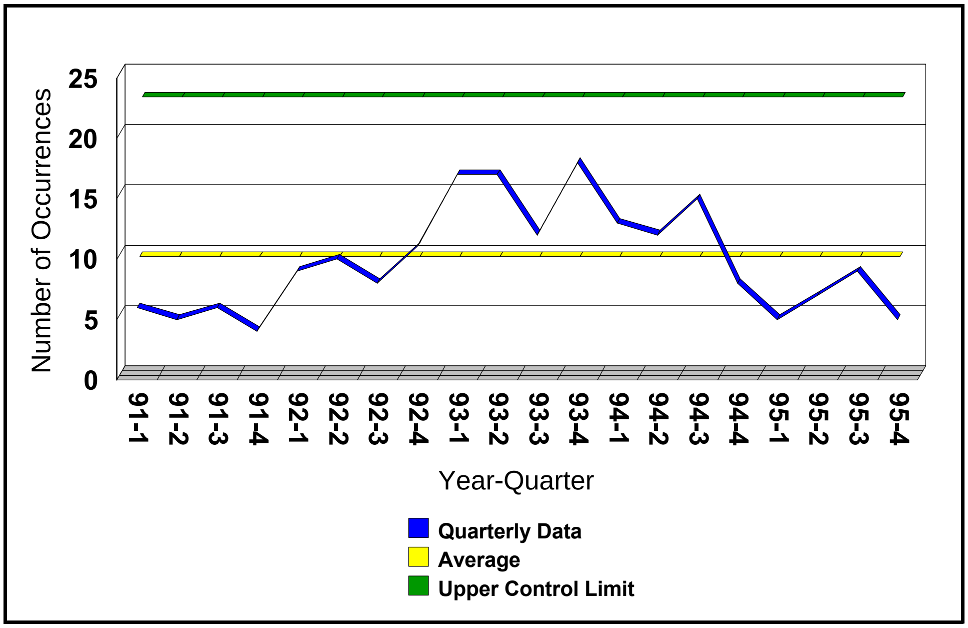


Figure 20 - Sample Control Chart

Histogram

A histogram is a bar chart that is designed to show a distribution of data. It is frequently used to display the frequency of events associated with measurements of time or cost. Data is grouped so that patterns of variability are easily identified.

Step-by-Step

TUTORIAL

Steps in Creating a Histogram

1. Assemble the data to be analyzed.
2. Determine appropriate scales for the charts.
3. Plot the data on the chart.
4. Analyze the chart for trends.

In the example shown in **Figure 21**, the number of construction related occurrences is distributed by year and quarter from 1991 through 1995. Distributions are provided for the total DOE and for the Chicago Field Office.

Evaluation of the DOE distribution shows the seasonal variation that would be expected for construction activities, with the lowest number of occurrences generally reported in the first quarter of each year. During 1991 and 1992, the number of occurrences was highest during the last quarter of each year. Further evaluation would be required to determine if this was due to weather, if it was a function of the level of construction (possibly related to the availability of new funding), or if it was a result of some other factor. The number of occurrences shows a decline during 1995.

The number of occurrences for Chicago also shows the seasonal variation. There is also a noticeable peak in 1993 that could be further investigated.

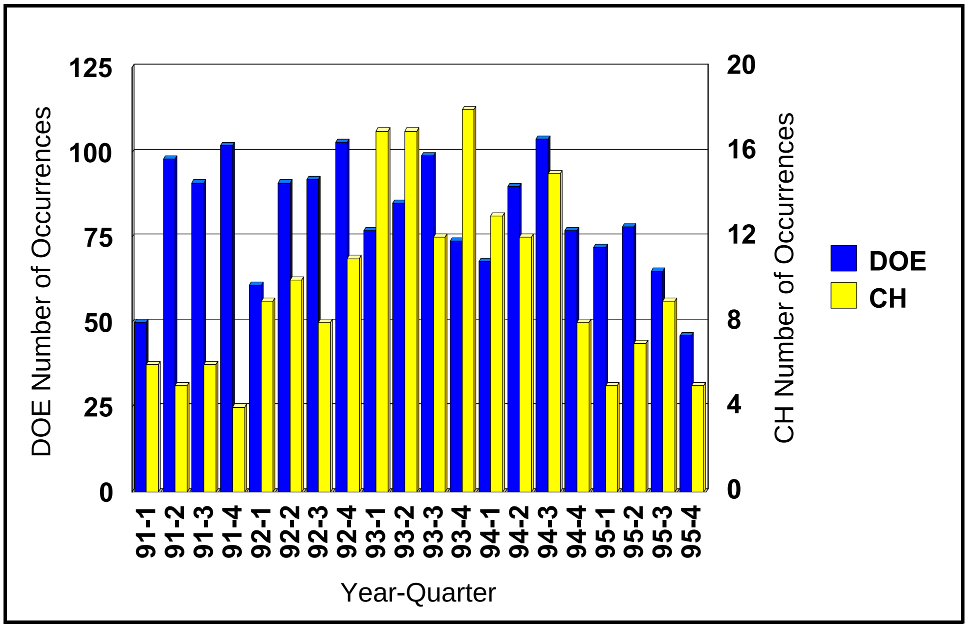


Figure 21 - Sample Histogram

Line Chart

A line chart is used to display the variation in a parameter over time.

Step-by-Step

TUTORIAL

Steps in Creating a Line Chart

1. Assemble the data to be analyzed.
2. Determine appropriate scales for the chart.
3. Plot the data on the chart.
4. Analyze the chart for trends.

Based on the bar chart of major contributors to the Chicago Field Office construction occurrence rate, line charts are prepared for BNL and NREL to evaluate their performance over the five year period from 1991 through 1995. This is shown below.

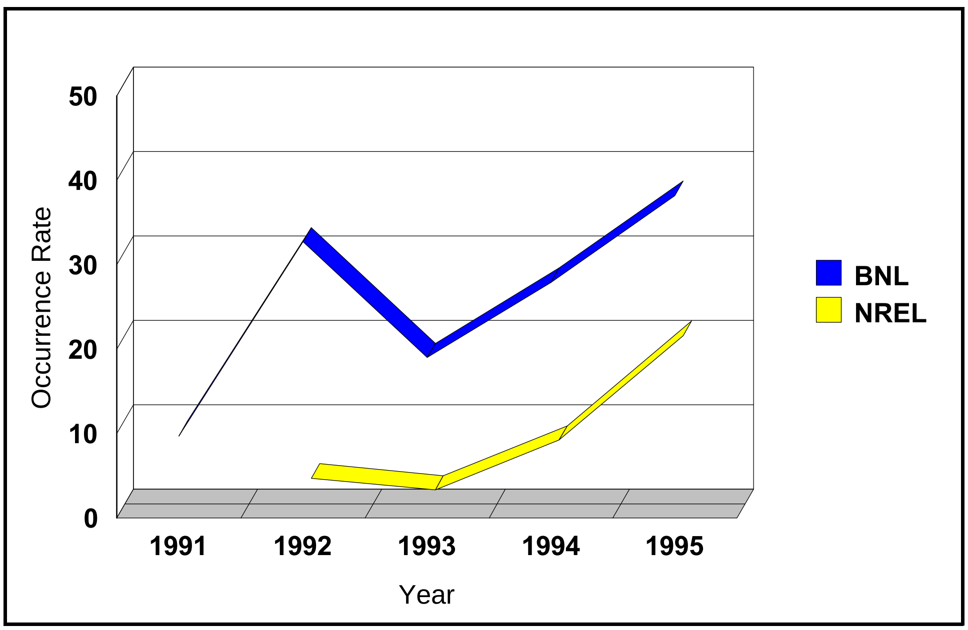


Figure 22 - Sample Line Chart

Pareto Analysis

Pareto Analysis is a simple method for separating the major issues associated with a problem from the minor ones. It is based on the relationship, observed in many fields, that a few issues are responsible for most of the problems. Pareto analysis can be used to prioritize issues and focus resources where they will have the most effect. It can help measure the impact of changes by comparing before and after conditions. It is a highly effective means of displaying the relative importance of issues.

TUTORIAL

Steps in Creating a Pareto Chart

1. Assemble the data to be analyzed.

2. Calculate the total of each item, the total for all items, and the percentage that each item represents of the total.

3. List the items in descending order, along with a cumulative percentage.

4. Draw a combined bar-line chart. The bars represent the total of each item, in descending order, while the line represents the cumulative percentage.

The Pareto analysis shown in **Figure 23** was used to rank Field organizations by total occurrence cost for all 1995 occurrence reports that specified an activity category of Construction. Costs were assigned to occurrences based on the values utilized for reports included under the value basis reporting criteria, i.e., $10,000 for each off-normal occurrence and $1,000,000 for each unusual occurrence. A value of $1,000,000 was also assigned to the single Emergency report included in this selection. The total cost was $15,480,000 for 13 unusual occurrences and 248 off-normal occurrences. This ranking was used in the Economic Matrix example in the Process Focus Tools section.

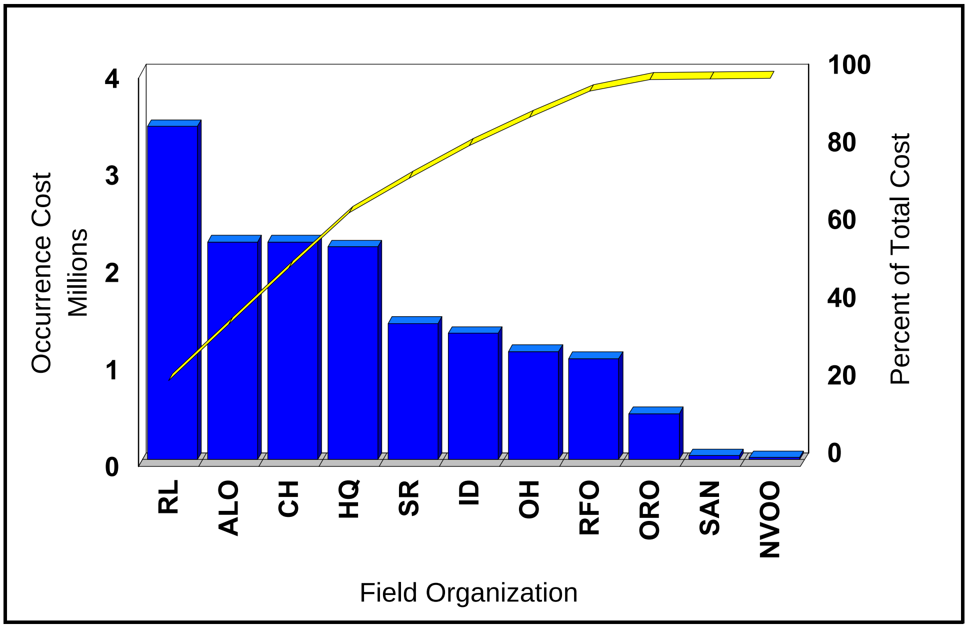


Figure 23 - Sample Pareto Analysis

NOTE For a small number of issues, the data for a Pareto analysis can be easily calculated manually. Where a large number of issues exist, for example, in comparing accident statistics for each contractor in DOE, the sorting and calculational features of a spreadsheet such as Excel are extremely useful.

Pie Chart

A pie chart is used to show the relationship of the components of a distribution within a set of data.

Step-by-Step

TUTORIAL

Steps in Creating a Pie Chart

1. Assemble the data to be analyzed.

2. Plot the data on the chart.

3. Analyze the chart for trends.

In **Figure 24**, the distribution of construction related occurrences by nature of occurrence is presented for the Chicago Field Office and the total DOE. Distributions are shown for both 1995 and for a reference period from 1991 through 1994.

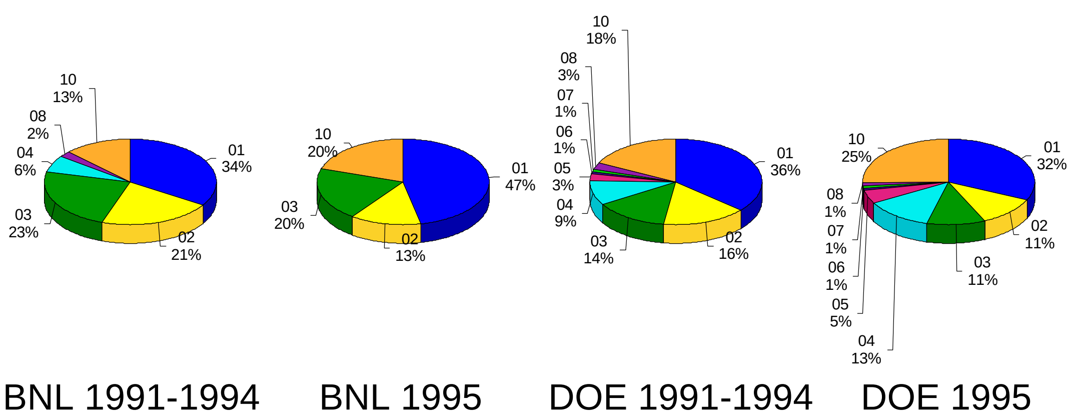


Figure 24 - Sample Pie Charts

Scatter Diagram

A scatter diagram helps determine if a relationship exists between two variables.

Step-by-Step

TUTORIAL

Steps in Creating a Scatter Diagram

1. Assemble the data to be analyzed.

2. Determine appropriate scales for the charts.

3. Plot the data on the chart.

4. Draw a best fit straight line for the data.

5. Analyze the degree to which the data fits the linear model.

In **Figure 25**, lost workday cases for the Chicago Field Office are analyzed to see if a correlation exists between the number of lost workdays per case and the age of the injured employee. This data comes from the CAIRS data base (CAIRS Field Office Codes 20 and 28). For each lost workday accident, the number of lost workdays is plotted against the age of the injured employee.

Although the data shows a fairly distinct upward trend that could indicate age as a factor in the extent of injury and/or recovery, few of the data points are close to the best fit line. A correlation is not proven. Note that the number of cases used in this example is quite limited due to the small amount of data available. A larger set of data would be required to prove or disprove an age correlation.

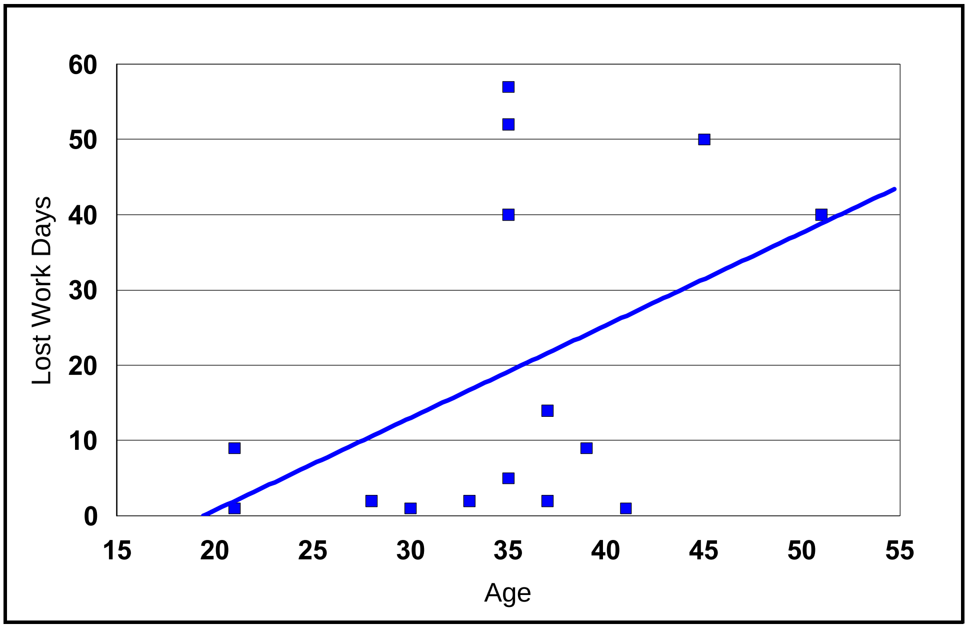


Figure 25 - Sample Scatter Diagram

Spider Diagram

The Spider Diagram, also known as a Radar Chart, provides easy visualization of the comparative rankings of several related issues. The Spider Diagram is especially useful when you want to visualize not only the relative performance of a number of issues, but also the relationship of each issue to some standard. If the outer limit of the chart is defined as the standard, e.g., a 100 percent rating on an evaluation, the position of each point with respect to the outer ring is a visual indication of performance.

Step-by-Step

TUTORIAL

Steps in Creating a Spider Diagram

1. Assemble the data to be analyzed.

2. Determine appropriate scales for the charts.

3. Plot the data on the chart.

4. Analyze the chart for trends.

Figure 26 shows the rate of construction related occurrence reports per 200,000 construction hours, as previously portrayed using the Bar Chart.

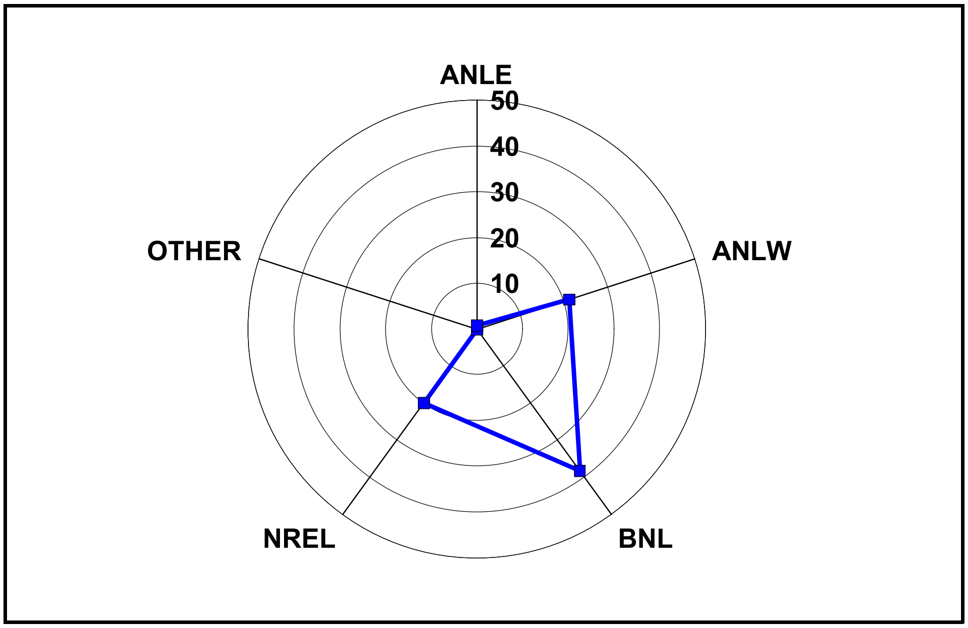


Figure 26 - Sample Spider Chart

Index